



SEMINARIO DE TÍTULO INGENIERO COMERCIAL
MENCIÓN ECONOMÍA

Universidad de Chile
Facultad de Economía y Negocios

**EL USO DE MATCHING PARA LA ESTIMACIÓN DE LA
ELASTICIDAD INTERGENERACIONAL DEL INGRESO***

Diego Polanco**
Profesor Guía: Javier Núñez

Santiago, 2012

*Agradezco la ayuda, las ideas y los comentarios hechos por Jean Paul Quinteros, Claudia Martinez, Juan Díaz, Nicolás Grau, Jorge Rivera, Julio Rodriguez y especialmente a mi profesor guía Javier Núñez, que hicieron posible la realización de esta tesis.

** dpolancon@fen.uchile.cl

Resumen

Esta investigación estudia la magnitud de la movilidad intergeneracional del ingreso para Chile, a través de una metodología de matching, nunca antes ocupada para estudios de este tipo. Ocupando la encuesta de Caracterización Socioeconómica Nacional en dos periodos distintos, 1992 y 2009, proponemos alternativas para dos limitaciones de la metodología que es ocupada usualmente por el concierto internacional, Variables Instrumentales en 2 Muestras: 1. imputar un ingreso para los padres de manera que éstos tengan una distribución bien comportada y 2. tener una metodología que permita solucionar para cualquier tipo de variables el problema de no tener bases intergeneracionales, lo que es frecuente en países en desarrollo como Chile. Las estimaciones para la elasticidad intergeneracional del ingreso se registraron entre 0.56-0.75, valores cercanos a los registrados a través de la metodología VI2M en estudios realizados anteriormente para nuestro país.

1. Introducción

La movilidad social se refiere al grado en el cual, dentro de una determinada estructura social, los individuos pueden cambiar de estatus social. Esto puede ser en el transcurso de su vida, lo cual es llamado **movilidad intrageneracional**, o a través de distintas generaciones, lo que es conocido como **movilidad intergeneracional**.

La **movilidad intergeneracional** es la forma en la cual los progenitores transmiten ventajas o desventajas, en una determinada sociedad, a sus hijos. Ésta tiene alcance en distintas variables socioeconómicas, tales como la categoría ocupacional, la escolaridad y el ingreso entre otras. Estudiar la movilidad intergeneracional tiene fundamento tanto por razones de eficiencia como por razones de economía normativa. Algunas de éstas son que:

- A un mayor grado de movilidad intergeneracional, es probable que se reduzcan las desigualdades económicas, promoviendo la justicia social y logrando una asignación de recursos más equitativa.
- Bajo el supuesto de que la distribución de los talentos en una sociedad es aleatoria, que exista movilidad intergeneracional es importante pues a mayores grados de movilidad social menos talentos se desperdician, lo cual podría tener consecuencias relevantes para el crecimiento económico.

En las últimas dos décadas, la literatura internacional ha realizado múltiples aportes para tratar de estimar el grado de movilidad intergeneracional de diversos países. Sin embargo, éstos estudios por lo general se han caracterizado por realizarse en países que tienen disponible bases de datos intergeneracionales.¹ Para realizar estudios como éstos en Chile, nos enfrentamos al problema de no tener bases de datos intergeneracionales con información de pares de progenitores e hijos/as.

A pesar de estos obstáculos, Núñez y Risco (2004) y Núñez y Miranda (2010) han realizado aportes estimando la elasticidad intergeneracional del ingreso entre padres e hijos siguiendo la metodología de Variables Instrumentales en 2 Muestras (VI2M) propuesta por Björklund y Jäntti (1997). Esta metodología ha sido utilizada ampliamente en el concierto internacional, lo cual permite poder comparar el grado de movilidad intergeneracional entre países. Sin embargo la metodología de VI2M, adolece de dos limitaciones:

¹Generalmente los países que tienen disponible bases intergeneracionales, son países de baja desigualdad y alta movilidad social, caso contrario al de nuestro país.

- Ésta solo puede realizar estimaciones de movilidad intergeneracional del ingreso. La metodología no puede rescatar información de los padres de otras variables para realizar otro tipo de estudios de movilidad intergeneracional. La movilidad intergeneracional en los años de escolaridad es un ejemplo.
- La distribución de los ingresos de los padres que se obtiene al implementar la metodología de VI2M, por construcción, no se encuentra bien comportada. Esto toma importancia cuando queremos estudiar la movilidad intergeneracional del ingreso entre las madres y sus hijos o hijas, debido a que de esta manera no se pueden implementar modelos que consideren sesgos de selección dados por la baja participación laboral femenina. Si bien Zenteno (2011) utiliza correcciones tipo Heckman descritas en Vella (1998) para tomar en cuenta una perspectiva de género considerando sesgos de selección, se restringe a estudiar la movilidad intergeneracional entre padres e hijas, sin considerar a las madres y no sorteja la limitación de que las distribuciones de los ingresos de los padres se encuentran por construcción mal comportadas.

La contribución que se trata de realizar con esta investigación a los estudios de movilidad intergeneracional, es tratar aflojar las restricciones que imponen estas dos limitantes. Para esto implementamos una imputación de los ingresos de los padres a través de un matching exacto, ocupando el algoritmo propuesto por Ñopo (2008).

El problema que tendremos para sacar conclusiones pertinentes a raíz de los resultados obtenidos a través de esta nueva metodología, es que no podrá ser comparada con las investigaciones realizadas internacionalmente. Sin embargo, al implementarse en la misma población que los estudios realizados por Núñez et. al. (2004) y (2010), sí podrá ser comparado con éstos, y es precisamente ésta la razón por la cual esta investigación se verá restringida solamente a la movilidad intergeneracional del ingreso. Para esto utilizaremos la Encuesta de Caracterización Socioeconómica Nacional.²

Esta investigación se estructurará de la siguiente manera: en primer lugar, se explicará la estrategia empírica y las metodologías econométricas usadas para nuestras estimaciones, es decir la de VI2M y la de matching exacto. A continuación revisaremos

²La Encuesta CASEN, o Encuesta de Caracterización Socioeconómica Nacional, es una encuesta a nivel nacional, regional y comunal, que realiza el gobierno de Chile desde el año 1985, con un periodicidad bianual y trianual. Para efectos de esta investigación se utilizaron las encuestas CASEN 1992 y CASEN 2009.

los datos utilizados, para luego revisar los resultados obtenidos. Finalmente se concluye.

2. Metodología

Para estimar el grado de movilidad intergeneracional, identificaremos la siguiente ecuación, con la que estimaremos la elasticidad intergeneracional del ingreso β_1 :

$$Y_i^h = \beta_0 + \beta_1 Y_i^p + \mu_i \quad (1)$$

Donde Y_i^h es el log ingreso permanente del hijo, Y_i^p es el log ingreso permanente del padre o la madre y μ_i un ruido blanco. Las conclusiones que nos entrega esta identificación pueden inferirse tomando casos extremos; si $\beta_1 = 1$, tendremos un bajo nivel de movilidad social, pues los ingresos de los hijos e hijas quedarían completamente determinados por los ingresos de las madres, si $\beta_1 = 0$ sería el caso contrario.

Por ende el grado de movilidad intergeneracional del ingreso, estará dado por $1 - \beta_1$

2.1. Variables Instrumentales en 2 Muestras

Björklund y Jäntii (1997) propusieron la metodología de Variables Instrumentales en 2 muestras, con tal de poder estimar el ingreso de los padres dado el reporte de los hijos, para luego estimar la elasticidad intergeneracional del ingreso. Por lo cual tenemos dos etapas de la estimación.

El principal problema al cual nos enfrentamos al identificar nuestro parámetro es el hecho de no poder observar el ingreso permanente de los individuos. Por ende, apegándonos a la teoría del ciclo de vida, utilizaremos en nuestras estimaciones, datos para los ingresos corrientes de los individuos en un período t .

Sabemos que los ingresos corrientes del padre provenientes del trabajo en el período t vendrán dados por la siguiente ecuación:

$$Y_{it}^p = Y_i^p + \alpha_0 \text{edad}_{it}^p + \alpha_1 \text{edad}_{it}^p{}^2 + \mu_{it} \quad (2)$$

Donde Y_{it}^p son los ingresos corrientes provenientes del trabajo del padre en el período t , Y_i^p es el ingreso permanente del padre y edad_{it}^p es la edad del padre i , en el período

t , con lo cual tenemos un modelo que considera el ciclo de vida.

Dado que no podemos observar el ingreso permanente de los individuos, podemos suponer que éste está dado por un vector de características socioeconómicas del individuo, es decir:

$$Y_i^p = Z_i\gamma + \nu_i \quad (3)$$

Por lo tanto nuestro modelo para el ingreso de los padres en el periodo t estará dado por:

$$Y_{it}^p = Z_i\gamma + \alpha_0 edad_{it}^p + \alpha_1 edad_{it}^{p^2} + \nu_i + \mu_{it} \quad (4)$$

$$Y_{it}^p = Z_i\gamma + \alpha_0 edad_{it}^p + \alpha_1 edad_{it}^{p^2} + \varepsilon_{it} \quad (5)$$

Debido a que no observamos los ingresos de las madres, ocupamos un vector Z de variables instrumentales con tal de estimar los ingresos de las madres y usarlos en la segunda etapa para estimar el ingreso permanente.

La metodología de Variables Instrumentales en 2 muestras, propone que dado que no observamos los ingresos en la muestra I , si es que tenemos otra muestra J que pertenece a la misma población de la muestra I , podemos estimar parámetros en J y ocuparlos en la muestra I con tal de estimar los ingresos del padre, es decir estimar el modelo de la ecuación (5) en la encuesta CASEN 1992

Teniendo aquellos parámetros estimados, podemos estimar los ingresos del padre para la encuesta Casen 2009 dado los parámetros estimados en la ecuación (5), como se muestra en la ecuación (6).

$$\widehat{Y}_{it}^p = Z_i\widehat{\gamma} + \widehat{\alpha}_0 edad_t + \widehat{\alpha}_1 edad_t^2 \quad (6)$$

Dado que estamos considerando el ciclo de vida en nuestras estimaciones, necesitamos estimar los ingresos en la misma etapa del ciclo de vida para todos los individuos, es decir ocupamos la variable $edad_t$ estandarizada para todos los individuos. La edad estandarizada que consideramos fue el promedio de edad de los “padres” de la encuesta CASEN 1992 que tenían hijos entre 6 y 18 años. La razón de esto es que consideramos que la transmisión intergeneracional del ingreso principalmente está dada por el nivel

de capital humano que los padres invierten en sus hijos, inversión que se daría en la educación básica y media, es decir entre los 6 y los 18 años de los hijos, excluyendo la educación superior por el hecho de que ésta puede ser financiada con créditos obtenidos en el mercado financiero. Esta edad estandarizada registrada en la CASEN de 1992 fue de 38 años para los padres.

Con esto tenemos una estimación de los ingresos de los padres según la educación y/u ocupación reportada por sus hijos en la encuesta CASEN 2009.

Luego podemos estimar nuestra ecuación de interés, de la cual podemos obtener la elasticidad del ingreso intergeneracional entre padres e hijos, la cual controlamos por el efecto del ciclo de vida de la siguiente manera:

$$Y_{it}^h = \beta_0 + \beta_1 \widehat{Y_{i(t=92)}^p} + \delta_0 edad_{it}^h + \delta_1 \widetilde{edad}_{it}^h{}^2 + \varepsilon_{it} \quad (7)$$

Es relevante considerar que en esta metodología podemos identificar dos sesgos. En primer lugar, dado que la estructura de la ecuación con la cual hemos identificado el ingreso permanente en relación con los ingresos corrientes, estos últimos indudablemente estarán correlacionados con el término de error ν_i , lo que nos presenta un problema de error de medición de libro de texto, lo que nos llevara a una estimación sesgada hacia abajo del estimador β_1 . Solon (1989) y (1992).

El segundo sesgo que podemos identificar es que si el instrumento Z escogido está correlacionado con sus ingresos, también es un posible predictor de los ingresos de los hijos. Dado esto, en la segunda etapa de nuestra metodología, la escolaridad y la ocupación fueron ocupadas para predecir el ingreso de los padres. Sin embargo, no están incluidas como regresores de los ingresos de los hijos, por lo que tendremos un problema de variables omitidas, lo que como el libro de texto nos diría, nos llevará a una estimación sesgada hacia arriba del parámetro de la elasticidad intergeneracional del ingreso, tal y como lo muestra Solon (1992).

El argumento de la literatura internacional para decir que la metodología VI2M es la adecuada para estimar la elasticidad intergeneracional del ingreso es el hecho de que los dos sesgos que se tienen por construcción de la estructura de la estimación se contraponen por lo cual debieran anularse entre sí entregando la estimación correcta de la elasticidad inetergeneracional del ingreso.

2.2. Matching Exacto

Rosebaum y Rubin (1983) realizaron un gran aporte a la investigación de diseños experimentales con la introducción del *propensity score* para ocupar el matching como una herramienta para identificar efectos causales. A partir de ahí la técnica de matching, a pesar de ser un instrumento estadístico, ha sido usada principalmente para estudios de evaluación de impacto. Sin embargo, este instrumental no ha sido exclusivamente ocupado para estos fines, de hecho Ñopo (2008) ocupa un matching exacto con tal de poder ofrecer una alternativa a Blinder-Oaxaca para la estimación de brechas salariales.

Precisamente la contribución de esta investigación está dada por la utilización de matching para otro tipo de estudios econométricos, en este caso para implementar una nueva metodología que permita hacer estudios de movilidad intergeneracional. Ésta consiste en un matching exacto en variables observables con tal de imputar el ingreso de un padre sintético a un hijo, dadas las características que éste reporta sobre su verdadero padre.

Para efecto de esta metodología utilizamos el mismo procedimiento utilizado por Ñopo (2008), pero implementado para pares de padres e hijos con efecto de estudiar movilidad intergeneracional. En este caso se tomarán individuos de una muestra, que en efecto de la investigación llamaremos hijos, y se les buscará un match con un padre sintético proveniente de otra muestra de la misma población, que tenga las mismas características observables que las que el hijo reporta sobre sus padres, pero de años anteriores. Se consideran para efectos de este matching sólo variables discretas y un matching exacto.

El algoritmo del matching exacto puede ser resumido de la siguiente manera:

- Paso 1: Seleccionar un hijo de la muestra (sin reemplazo).
- Paso 2: Seleccionar todos los padres de la muestra de años anteriores que tengan exactamente las mismas características observables.
- Paso 3: Construir un padre sintético cuyo ingreso es el promedio o la mediana del ingreso de todos los individuos seleccionados en el paso 2, y hacer un match con el hijo en cuestión.
- Paso 4: Poner todas las observaciones de ambos individuos (el hijo y su padre sintético) en la nueva muestra de individuos con su match respectivo.

- Repetir el paso 1 hasta el 4, hasta encontrar un padre sintético para todo hijo de la muestra.

De esta manera tenemos una imputación del ingreso del padre para los hijos de la muestra, lo que nos permite tener los datos para nuestra ya conocida identificación econométrica:

$$Y_i^h = \beta_0 + \beta_1 \overline{Y_i^p} + \mu_i \quad (8)$$

Donde $\overline{Y_i^p}$ es el ingreso del padre imputado para el individuo “ i ” a través del algoritmo de matching.

Previo a correr la regresión, eliminamos todas las observaciones que no se encuentran en el soporte común.

Es importante tomar en cuenta, que en este caso no todo individuo “ i ” tiene el mismo número de padres posibles, es decir cada individuo “ i ” tiene n_i padres. Por lo tanto, por ley de los grandes números, debiera darse que la imputación a través del matching exacto será más precisa a medida que el n_i sea más grande. Por ello también realizamos estimaciones con un ponderador que castiga el ingreso de los padres imputados con menor cantidad de padres posibles, es decir con menor n_i .

Para efecto de nuestras estimaciones utilizamos 2 ponderadores para las regresiones ponderadas, donde n_i es el total de padres que cumplen con el matching exacto, y N es el total de padres posibles.

- Uno lineal, calculado como $\lambda_i = \frac{n_i}{N}$
- Uno no-lineal, calculado de la siguiente manera: $\lambda_i = \frac{\sqrt{n_i}}{N}$

La idea de tener un ponderador no-lineal, es que el aporte marginal de tener una mayor cantidad de padres que cumplan con el matching exacto sea decreciente, debido a que al tener una mayor cantidad de observaciones, tener una adicional no aporta tanto a que la media o la mediana del ingreso sea una imputación más precisa del ingreso de su padre verdadero.

Es importante tomar en cuenta que los ingresos que están siendo imputados, son la mediana o media de ingresos corrientes no de ingresos permanentes, por lo cual deberían estar correlacionados con el término de error, lo cual nos lleva, al igual que en

la metodología de VI2M, a un sesgo a la baja de nuestro estimador del parámetro β_1 debido al error de medición de libro de texto.

A diferencia de la metodología VI2M, en esta metodología no tenemos un sesgo a la baja debido a que al no ocupar la educación como variable instrumental para la estimación de los ingresos de los padres, no se presentara el problema de variables omitidas como sí ocurre en la metodología VI2M.

Sin embargo, el sesgo a la baja del estimador a través de esta metodología será atenuado siempre que cada hijo tenga un padre sintético creado a partir de mayor cantidad de individuos, es decir mayor n_i , dado que al provenir la imputación de una mediana o una media, el ingreso se ve atenuado cuando recibe el shock al alza o a la baja que tienen los ingresos corrientes en relación al ingreso permanente.

Para tomar en cuenta el efecto del ciclo de vida de los hijos, también se controló por la edad y la edad al cuadrado de éstos, corriendo la siguiente regresión:

$$Y_i^h = \beta_0 + \beta_1 \overline{Y_i^p} + \delta_0 edad + \delta_1 edad^2 + \mu_i \quad (9)$$

Finalmente debemos reconocer que esta metodología, como se presenta en esta investigación, carece de formalidad matemática. Esto no permite identificar la varianza de nuestras estimaciones y por ende no podemos establecer cuál es el intervalo de confianza de éstas. Esto es tarea para futuras investigaciones.

3. Datos

Como hemos mencionado, esta investigación se ha realizado con datos de la Encuesta de Caracterización Socioeconómica (CASEN) de dos distintos periodos. Ésta es representativa a nivel regional y nacional, y realizada periódicamente desde su primera versión en el año 1987. En esta investigación se utilizaron dos encuestas, del año 1992 y 2009.³ La encuesta del año 1992 cuenta aproximadamente con una muestra de 36.000 hogares, mientras que la encuesta del año 2009 cuenta con aproximadamente 70.000. La encuesta CASEN provee información socioeconómica estandarizada sobre los jefes de hogares y otros miembros adultos del hogar, incluyendo genero, edad, establecimiento educacional, situación laboral, ocupación, ingresos corrientes provenientes del trabajo,

³Se ocuparon dos encuestas CASEN que tuvieran una diferencia de años tal, para que los individuos de la primera encuesta tuvieran la edad suficiente ser padres de los individuos de la segunda

entre otros, tanto para áreas urbanas como para las rurales.

Para las estimaciones tanto de la primera etapa de la metodología de variables instrumentales en dos muestras, como para la metodología de matching se tomaron en cuenta como padres a individuos entre 15 y 55 años de la encuesta CASEN 1992, los cuales se registraron en un total 19,089 individuos.

En esta investigación nos restringimos a estimar la elasticidad intergeneracional del ingreso sólo para pares de padres e hijos, debido a que al considerar sólo hombres las estimaciones no se ven sujetas a problemas de sesgo de selección por la falta de participación en el mercado laboral, como sí pasa con las mujeres. Una segunda justificación es que la evidencia internacional también lo hace de esta manera lo cual permite tener una medida comparable. Núñez et. al., también lo hacen de esta manera, por lo cual para poder comparar nuestras estimaciones con sus investigaciones, es necesario que tomemos esta restricción de nuestra investigación.

Para las estimaciones de la segunda etapa de VI2M y para la metodología de matching se consideraron como hijos sólo a hombres jefes de hogar o cónyuges debido a que son los únicos que responden la sección autobiográfica de la encuesta CASEN, en la cual se recoge la información sobre los padres. Se eliminaron todas las observaciones de los individuos que no recordaban la ocupación o el nivel educacional de su padre o madre. Se tomaron en cuenta sólo individuos entre 25 y 40 años que reportaran ingresos, para que las estimaciones no fueran contaminadas por posibles sesgos dados por el ciclo de vida⁴ registrándose un total de 6,486 individuos.

En ninguna de las muestras consideramos a quienes no reportan ingresos, debido a que la encuesta CASEN es un corte transversal y a nosotros nos interesa la relación del ingreso permanente entre progenitores y sus hijos, por lo cual que un individuo no participara en el mercado laboral en determinado momento del tiempo no nos da mayor información en el contexto intergeneracional.

⁴Haider y Solon (2006) realizan estimaciones considerando que la relación del ingreso con el ciclo de vida no está dada por un paseo aleatorio, concluyendo que al estimar la movilidad intergeneracional del ingreso puede estar sesgada en las primeras etapas del ciclo de vida, por lo cual no consideramos individuos menores de 25 años.

4. Resultados

En la Tabla 1, presentamos el valor de las estimaciones de la elasticidad intergeneracional del ingreso a través de la metodología VI2M, es decir el valor del estimador β_1 . Las estimaciones presentadas en la columna 2, fueron obtenidas estimando la primera etapa con variables de ocupación y de educación, en cambio las estimaciones presentadas en la columna 3 fueron obtenidas estimando la primera etapa solamente con variables de educación.

Tabla 1. Estimación Elasticidad Intergeneracional del Ingreso. Metodología VI2M

	Ingreso del padre estimado por educación y ocupación	Ingreso del padre estimado por educación
<i>y</i> padre	0.496***	0.636 ***
<i>edad</i>	0.1035***	0.1002***
<i>edad</i> ²	-.0013355***	-.001315***
Constante	1.3609**	2.399***
<i>R</i> ²	0.091	0.0906
Observaciones de la 1era etapa: 19,089		
Observaciones de la 2da etapa: 6,468		
Desviación Estándar en Paréntesis		
*** p<0.01, ** p<0.05, * p<0.1		

Como podemos ver en la Tabla 1, la elasticidad intergeneracional del ingreso entre padres e hijos está acotada en un rango de 0,49–0,64. Estas estimaciones nos dan un valor para β_1 que es alto al compararlo con otros países, esto es similar a las estimaciones las realizadas por Núñez et. al (2004) y (2010) que se presentan en la Tabla B del apéndice.

Los resultados deben haber variado entre estas estimaciones y las de Núñez debido a la utilización de distintas encuestas en la primera etapa y por la muestra ocupada, dado que la utilizada en esta investigación se restringió solo a individuos que hayan contestado la encuesta autobiográfica de manera completa con tal de que pudiera tenerse información suficiente para poder realizar el matching propuesto con mayor cantidad de variables.

En la Tabla 2, presentamos un test ANOVA⁵ de diferencia de medias, con tal de tratar de identificar cuál es la especificación más apropiada de las variables que debemos ocupar en nuestro matching. La idea de esto es ver si es que los distintos grupos que se hacen a través de las interacciones de todas las variables cualitativas tienen diferencias

⁵Análisis de Varianza

significativas en el ingreso, que es precisamente lo que queremos identificar en nuestro modelo econométrico. El criterio utilizado para esto es uno de robustez, considerando el R^2 ajustado como el que nos entrega mayor información sobre la variable dependiente que en este caso es el ingreso por hora.

Dentro de las variables consideradas para implementar el matching exacto, consideramos variables socioeconómicas como la educación, la categoría ocupacional y el número de habitantes del hogar. También consideramos la educación del cónyuge, debido a que puede darse que cierto porcentaje de la movilidad intergeneracional del ingreso esté explicada por el hecho de que las personas de altos ingresos se casen con personas de altos ingresos, esto es modelado por Ermisch, Francesconi y Siedler (2006). También incluimos variables de características geográficas como lo es la zona geográfica que indica si el individuo vive en zona rural o urbana, y una variable de región. Esta última, la incluimos bajo el supuesto de que los hijos considerados en la muestra de la encuesta CASEN 2009 no se cambiaron de la región en la cual vivieron con sus padres cuando tenían menos de 15 años. Esto lo sustentamos con las conclusiones de González y Rodríguez (2006), en el cual muestran que la tasa de entrada y salida de habitantes por región en Chile es considerablemente baja, por lo cual el supuesto establecido es razonable.

Tabla 2. Test ANOVA para diferencia de medias

H_0 : La media de los grupos son iguales entre si	
Variabes	R^2 Adj
Educación Ocupación	0.31
Educación Ocupación Zona Geográfica	0.31
Educación Ocupación Educación de la Madre	0.34
Región Educación Ocupación Zona Geográfica	0.40
Educación Ocupación Educación de la Madre Zona Geográfica	0.34
Región Educación Ocupación Educación de la Madre	0.43
Región Educación Educación de la Madre Zona Geográfica	0.21
Región Educación Ocupación Educación de la Madre Zona Geográfica	0.47
N de Habitantes en el Hogar Educación Ocupación Educación de la Madre Zona Geográfica	0.35
Para todas las especificaciones el test ANOVA resulto significativo al 90 %, 95 % y 99 % de confianza	

Como puede apreciarse, todas las especificaciones son estadísticamente significativas al 99 % de confianza. Utilizando el criterio de robustez a través del R^2 ajustado, el modelo más apropiado para implementar la metodología de matching exacto es la que considera las variables región, educación, ocupación, educación de la madre y zona geográfica.

A continuación se presenta en la Tabla 3, las estimaciones a través de la metodología de matching exacto, en la cual se ocupó la especificación más robusta, la cual considera las variables región, educación, educación de la madre, ocupación y zona geográfica. En la

Tabla 4, se presentan las mismas estimaciones utilizando la misma especificación, pero sin la variable ocupación con el fin de poder comparar con la metodología de VI2M, en la cual se implementa la primera etapa con variables educacionales y ocupacionales y también solamente con variables educacionales.⁶

Como lo explicamos anteriormente, al imputar el ingreso de los padres, lo hicimos a través de una media y también de una mediana. Esta última fue considerada porque al ser ingresos los que se están imputando, es frecuente la existencia de outlayers, los cuales ensucian la información que esta entregándonos el valor medio del ingreso de todos los padres que cumplen con el matching exacto, lo que no ocurre con el valor mediano.

A pesar de que se presentan los resultados para las estimaciones sin ponderadores, éstos no los consideramos como una estimación verídica de la elasticidad intergeneracional del ingreso. Esto es debido a que no podemos considerar como igualmente precisa la imputación del ingreso de un padre sintético hecha con 1 observación que la imputación del ingreso de un padre sintético construida con 100 observaciones. Evidentemente, que por ley de los grandes números, la segunda será más precisa.

Tabla 3. Estimación de Elasticidad Intergeneracional del ingreso. Metodología Matching Exacto

Variables utilizadas en el matching: región, educación, ocupación, educación de la madre, zona geográfica						
Promedio						
	Sin Ponderadores		Ponderador Lineal		Ponderador no-lineal	
<i>y</i> padre	0.408***	0.414***	0.526***	0.541***	0.479***	0.489***
<i>edad</i>		0.0921***		0.133***		0.107***
<i>edad</i> ²		-0.000121		-0.00183***		-0.00145***
Constante	4.903***	3.160***	4.171***	1.730***	4.460***	2.473***
<i>R</i> ²	0.095	0.101	0.105	0.112	0.103	0.108
Mediana						
	Sin Ponderadores		Ponderador Lineal		Ponderador no-lineal	
<i>y</i> padre	0.444***	0.450***	0.624***	0.641***	0.544***	0.555***
<i>edad</i>		0.0944***		0.131***		0.107***
<i>edad</i> ²		-0.00125***		-0.00181***		-0.00145***
Constante	4.781***	3.005***	3.726***	1.309**	4.187***	2.214***
<i>R</i> ²	0.096	0.102	0.105	0.112	0.102	0.107

6,287 Observaciones
181 Observaciones eliminadas por no estar en el soporte común
Desviaciones Estándar en Paréntesis
*** p<0.01, ** p<0.05, * p<0.1

⁶En la Tabla C y D del apéndice se muestran tablas con algunas estadísticas descriptivas sobre los padres posibles para ambas especificaciones

Tabla 4. Estimación de Elasticidad Intergeneracional del ingreso. Metodología Matching Exacto

Variables utilizadas en el matching: región, educación, educación de la madre, zona geográfica						
Promedio						
	Sin Ponderadores		Ponderador Lineal		Ponderador no-lineal	
<i>y</i> padre	0.493***	0.508***	0.564***	0.588***	0.532***	0.551***
<i>edad</i>		0.0941***		0.140***		0.121***
<i>edad</i> ²		-0.00120**		-0.00189***		-0.00161***
Constante	4.334***	2.447***	3.833***	1.143**	4.058***	1.728***
<i>R</i> ²	0.118	0.126	0.128	0.138	0.122	0.130
Mediana						
	Sin Ponderadores		Ponderador Lineal		Ponderador no-lineal	
<i>y</i> padre	0.603***	0.617***	0.717***	0.746***	0.657***	0.676***
<i>edad</i>		0.0866***		0.137***		0.116***
<i>edad</i> ²		-0.00109**		-0.00184***		-0.00155***
Constante	3.865***	2.129***	3.150***	0.500	3.526***	1.290**
<i>R</i> ²	0.118	0.126	0.128	0.138	0.122	0.130

6,447 Observaciones

21 Observaciones eliminadas por no estar en el soporte común

Desviaciones Estándar en Paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Como se puede ver, tanto en la estimación en la cual se usó el promedio como en la que se usó la mediana, el ponderador lineal entregó estimaciones de mayor magnitud que las estimaciones en las cuales usamos el ponderador no-lineal. Esto nos entregó los resultados que esperábamos dado que el ponderador lineal castiga el ingreso de los padres sintéticos construidos con menor cantidad de observaciones, al igual que el ponderador no-lineal. Sin embargo, este último castiga de manera creciente pero a tasas decrecientes, lo cual es consistente con la intuición de que si bien una imputación a través de una media o mediana es menos precisa cuando se realiza con una menor cantidad de observaciones, el aporte marginal de cada observación a la precisión es menor. Esto puede llevarnos a la conclusión de que el ponderador lineal sobrestima el parámetro identificado.

Sin embargo, esta sobrestimación producida por el ponderador lineal podría resultar beneficiosa para sacar algunas conclusiones. Si consideramos el sesgo a la baja que explicamos en la sección de la metodología, que se produce por utilizar ingresos corrientes y no permanentes, puede que éste se anule con la sobrestimación producida al ocupar el ponderador lineal, en vez del ponderador no-lineal. Este es el mismo argumento, en el cual se sostiene que la estimación a través de VI2M es precisa para la estimación de la elasticidad intergeneracional del ingreso.

Apegándonos a este argumento, podemos decir que el valor de la elasticidad intergeneracional del ingreso estará acotada entre los valores 0.62-0.72, y entre los valores 0.65-0.75, cuando se controla por el efecto del ciclo de vida de los hijos.

En la sección de metodología expusimos que el efecto a la baja por considerar ingresos corrientes en vez de ingresos permanentes podría estar aplacado por el hecho de que la metodología de matching exacto construye padres sintéticos a través de promedios o medianas, funciones que podrían absorber el shock que tienen los ingresos corrientes con respecto a los ingresos permanentes.

Considerando este argumento en vez del expuesto anteriormente, el valor de la elasticidad intergeneracional del ingreso estará acotada entre los valores 0.54-0.66, y entre los valores 0.56-0.68, cuando se controla por el efecto del ciclo de vida de los hijos.

Debe tenerse en cuenta que si bien todos los estimadores de la elasticidad intergeneracional del ingreso son significativos a un 99% de confianza, expusimos en la sección de la metodología de matching exacto que está no ha sido formalizada matemáticamente, por lo cual no conocemos el verdadero valor de la varianza. Una posible solución para este problema, que queda pendiente para futuras investigaciones, es la que utilizan Núñez y Miranda (2010), en el cual empena la corrección para la desviación estándar de Murhpy y Topel (1985).

Como mencionamos en la introducción, efectivamente la metodología de variables instrumentales en dos muestras imputa un ingreso estimado de los padres, que tiene una distribución sumamente mal comportada, lo cual puede verse en el apéndice donde mostramos la distribución de nuestra imputación por VI2M en la figura 1.

A través de la metodología de matching exacto, podemos imputar un ingreso para los padres sintéticos que tiene una distribución mucho mejor comportada que la obtenida a través de VI2M. Como puede observarse en las figuras 2, 3 y 4 del apéndice, la distribución mejora a medida que hacemos más restrictivo el matching cuando le agregamos más variables, de hecho puede verse que éstas se parecen muchísimo más a la figura 5, en la cual se muestra la distribución de ingreso por hora de la CASEN 1992, que la figura 1, en la cual se muestra la distribución del ingreso imputado a través de la metodología VI2M.

El mejor comportamiento de la distribución del ingreso a través de la metodología de VI2M ocurre bajo el costo de perder soporte común, lo cual nos hace perder precisión de nuestros estimadores considerando que se estima a través de MCO. Podría decirse que se establece un trade-off entre dos elementos que son deseables para nuestras estimaciones.

5. Conclusiones

En esta investigación hemos elaborado una nueva metodología a través de matching exacto, para estimar la elasticidad intergeneracional del ingreso. Siguiendo esta metodología estimamos la elasticidad intergeneracional del ingreso entre 0.56-0.75, dependiendo de si se controla o no por ciclo de vida, y cual es la especificación del matching que se utilizó para imputar el ingreso de los padres sintéticos.

Estos valores son sumamente consistentes con las estimaciones realizadas por Núñez y Miranda (2010), que se registran entre 0.57-0.74. Dado esto, podemos concluir que la metodología es válida para realizar estimaciones de movilidad intergeneracional del ingreso, lo cual nos da pie para pensar que podremos extenderla para realizar estudios de movilidad intergeneracional de otro tipo de características socioeconómicas que no se han realizado en nuestro país. El estimar a través de otra metodología y obtener resultados mucho más altos que países desarrollados y más equitativos, confirma las conclusiones de Núñez y Miranda (2010), sobre que Chile es un país con baja movilidad social, relativo al resto del mundo.

Lamentablemente, esta investigación carece de formalidad matemática lo cual no nos permite identificar la magnitud de la varianza de nuestras estimaciones. Esto no nos permite ver el verdadero valor de los intervalos de confianza y por esto no podemos afirmar con certeza que los estimadores sean estadísticamente significativos al mismo intervalo de confianza que obtenemos al estimar por mínimos cuadrados ordinarios.

Esta metodología logra sortear dos limitaciones que identificamos en la metodología de Variables Instrumentales en 2 Muestras:

- Poder imputar datos que no sean solamente de ingreso, lo que no fue implementado en esta tesis pero que puede llevarse a cabo sin mayores problemas con el algoritmo de Ñopo (2008).
- Tal y como lo mostramos en los histogramas de los ingresos que hemos imputado a través de la metodología de matching exacto, conseguimos una distribución para el ingreso de los padres sintéticos mucho más parecida a una distribución del ingreso de los padres en la CASEN 1992 de la que se logra conseguir a través de la metodología de VI2M.

Queda pendiente para futuras investigaciones poder utilizar para modelos más complejos, las soluciones que propone la metodología empleada para los problemas de falta de

datos de ciertas variables socioeconómicas y la mejora en la distribución de los datos de ingreso que se puede obtener de esta manera. A su vez, también queda pendiente formalizar matemáticamente la metodología con tal de poder identificar la varianza y, por ende, los intervalos de confianza de nuestras estimaciones.

Apéndice

Tabla A. Estimación Elasticidad Intergeneracional por metodología VI2M

	Ingreso del padre estimado por educación y ocupación	Ingreso del padre estimado por educación
<i>edad</i>	0.0316*** (0.00470)	0.0411*** (0.00505)
<i>edad</i> ²	-0.000263*** (6.11e-05)	-0.000326*** (6.57e-05)
Básica Completa	0.220*** (0.0123)	0.217*** (0.0133)
Media Completa	0.656*** (0.0140)	0.684*** (0.0148)
Técnica Superior Completa	1.050*** (0.0409)	1.049*** (0.0440)
Universitaria Completa	1.621*** (0.0215)	1.655*** (0.0229)
Cuenta Propia	-0.760*** (0.0271)	
Trabajador Dependiente	-1.186*** (0.0256)	
Fuerzas Armadas	-0.927*** (0.0484)	
Servicios Domésticos	-1.587*** (0.169)	
Constante	5.920*** (0.0910)	4.614*** (0.0933)
<i>R</i> ²	0.380	0.280

Observaciones 19,089

Desviación Estándar en Paréntesis

*** p<0.01, ** p<0.05, * p<0.1

Tabla B. Evidencia de Elasticidad Intergeneracional
del Ingreso para Chile

Estudio	Tramo Etario	β_1
Núñez y Risco (2004)	23 – 55	0,58
Núñez y Miranda (2010)	25 – 40	0,57 – 0,74
Núñez y Miranda (2010)	25 – 30	0,45 – 0,72
Núñez y Miranda (2010)	31 – 35	0,59 – 0,73
Núñez y Miranda (2010)	36 – 40	0,66 – 0,79

Tabla C. Resumen de Frecuencias

Especificación del Matching: Región Educación Educación de la Madre Ocupación Zona Geográfica			
	N Padres	Ponderador Lineal	Ponderador No-Lineal
Max	838	0.0321294	0.0011099
Min	1	0.0000383	0.0000383
Promedio	16.89	0.000647672	0.00011038

Tabla D. Resumen de Frecuencias

Especificación del Matching: Región Educación Educación de la Madre Zona Geográfica			
	N Padres	Ponderador Lineal	Ponderador No-Lineal
Máximo	1147	0.0439767	0.0012985
Mínimo	1	0.0000383	0.0000383
Promedio	38.10	0.00146084	0.00016737

Figura 1. Histograma log Y por hora Metodología VI2M
 Ingresos Estimados a través de la primera etapa de VI2M

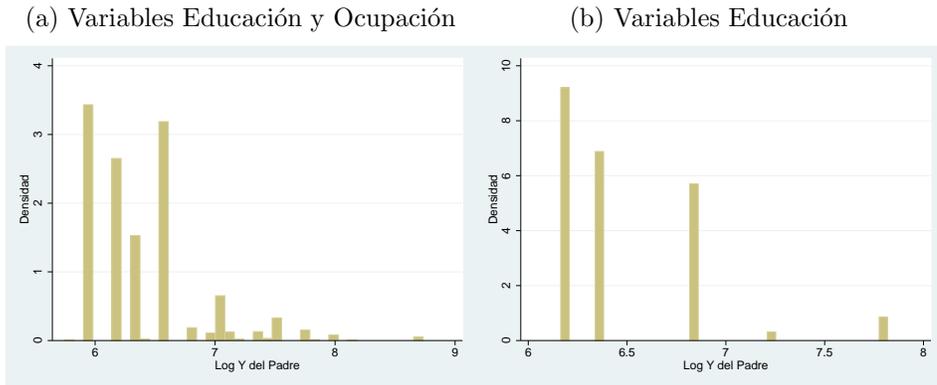


Figura 2. Histograma log Y por hora metodología Matching Exacto
 Especificación: región, educación, educación de la madre, zona geográfica
 Observaciones en el soporte común 6,447

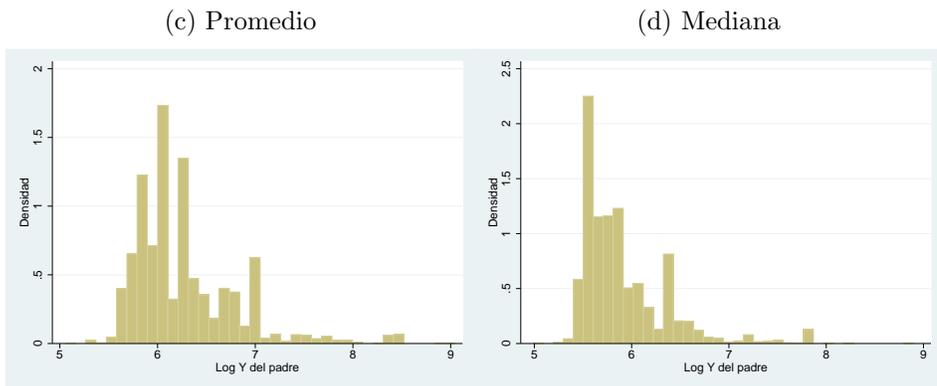


Figura 3. Histograma log Y por hora Metodología Matching Exacto
 Especificación: región, educación, ocupación, educación de la madre, zona geográfica
 Observaciones en el soporte común 6,287

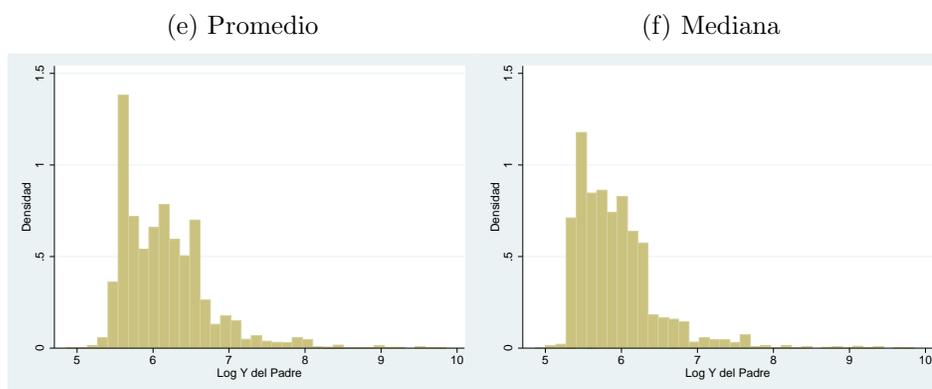


Figura 4. Histograma log Y por hora Metodología Matching Exacto
 Especificación: región, educación, ocupación, educación de la madre, zona geográfica,
 N de habt del hogar

Observaciones en el soporte común 4,003

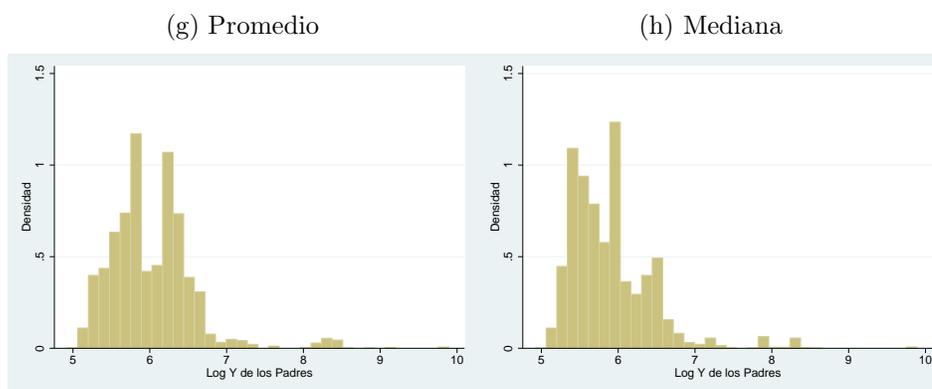
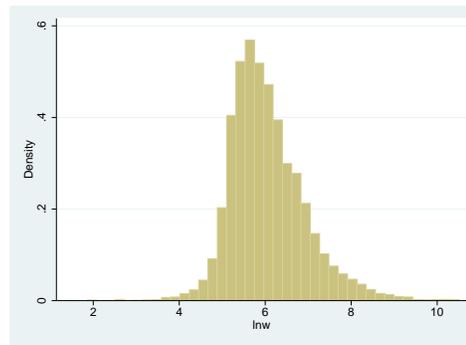


Figura 5. Histograma log Y por hora CASEN 1992



Referencias

- Angrist, J. y Krueger, A. (1992), "The effect of age at school entry on educational attainment: An application of instrumental variables with moments from two samples", *Journal of the American Statistical Association* 87(418), pp. 328-336.
- Becker, G. y N. Tomes (1979), "An equilibrium theory of the distribution of income and intergenerational mobility", *Journal of Political Economy* 87(6), pp. 1153-1189.
- Björklund, A. y Jäntti, M. (1997) "Intergenerational Income Mobility in Sweden Compared to the United States". *The American Economic Review*, Vol. 87, No. 5 (Dec., 1997), pp. 1009-1018
- d'Addio, A. (2007) "Intergenerational Transmission of Disadvantage: Mobility or Immobility across Generations? A Review of the Evidence for OECD Countries". *OECD Social, Employment and Migration Working Papers*.
- Ermisch, J. Francesconi, M. y Siedler, T. (2006) "Intergenerational Mobility and Marital Sorting". *The Economic Journal*, 116 (July), 659-679.
- Haider, S. y Solon, G. (2006). "Life-Cycle Variation in the Association Between Current and Lifetime Earnings". *NBER*. Working Paper No 11943.
- Minicozzi, A. (2002) "Estimation of sons's intergenerational earnings mobility in the presence of censoring". *Journal of Applied Econometrics*. 18: 291-314 (2003).
- Murphy, K. y Topel, R. (1985) "Estimation and inference in two-step econometric models". *Journal of Business and Economic Statistics* 3(4), pp. 370-379.
- Neilson, C. Contreras, D. Cooper, R. y Hermann, J. (2008) "Dinamic of Poverty in Chile". *Journal of Latin American Studies* (2008), 40 : pp 251-273
- Núñez, J. y Risco, C. (2004) "Movilidad Intergeneracional del Ingreso en un país en desarrollo: el caso de Chile". Documento de Trabajo N 210, Departamento de Economía. Universidad de Chile.
- Núñez, J y Perez, G. (2007) "Dime cómo te llamas y te diré quién eres: la Ascendencia como mecanismo de diferenciación social en Chile". Documento de Trabajo N 269.

Departamento de Economía. Universidad de Chile.

Núñez, J. y Miranda, L. (2010) “Intergenerational Income Mobility in a Less- Developed, High-Inequality Context: The Case of Chile” *The B.E. Journal of Economic Analysis and Policy*: Vol. 10: Iss. 1, Article 33.

Ñopo, H. (2008) “Matching as a Tool to Decompose Wage Gaps” *The Review of Economics and Statistics*, May 2008, 90(2): 290-299.

Rosenbaum, P. y Rubin, D. (1983) “The Central Role of the Propensity Score in Observational Studies for Causal Effects,” *Biometrika* 70:1. 41-55.

Solon, G. (1989) “Biases in the Estimation of Intergenerational Earnings Correlations”, *The Review of Economics and Statistics*, Vol. 71, No 1. (Feb 1989), 172-174.

Solon, G. (1992) “Intergenerational income mobility in the United States”, *American Economic Review* 82(3), pp. 393-408.

Solon, G. (1999) “Intergenerational Mobility in the Labor Market”. *Handbook of Labor Economics, Volume 3*, Edited by O. Ashenfelter and D. Card

Vella, F. (1998): “Estimating models with sample selection bias: A survey” *Journal of Human Resources* 32(1), pp 127-169.

Zenteno, I. (2011) “Movilidad Intergeneracional de los Ingresos en Chile: Una perspectiva de Género”. Seminario para optar al título de Ingeniero Comercial, Mención Economía. Universidad de Chile.